THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 48, pp. 34736–34745, November 29, 2013

Full Implementation of the Genetic Code by Tryptophanyl-tRNA Synthetase Requires Intermodular Coupling*

Received for publication, August 19, 2013, and in revised form, October 10, 2013 Published, JBC Papers in Press, October 20, 2013, DOI 10.1074/jbc.M113.510958

Li Li and Charles W. Carter, Jr.[1]

From the Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7260

Background: Aminoacyl-tRNA synthetases evolved from Urzymes with reduced amino acid specificity.
Results: Independent restoration of either the catalytic insertion domain or the anticodon-binding domain greatly reduces both amino acid specificity and tRNA aminoacylation.
Conclusion: Amino acid selectivity and tRNA acylation require interdomain cooperativity.
Significance: Independent recruitment of either module would have significantly reduced evolutionary fitness as Class I aaRS evolved.

Tryptophanyl-tRNA Synthetase (TrpRS) Urzyme (fragments A and C), a 130-residue construct containing only secondary structures positioning the HIGH and KMSKS active site signatures and the specificity helix, accelerates $tRNA^{Trp}$ aminoacylation with ~10-fold specificity toward tryptophan, relative to structurally related tyrosine. We proposed that including the 76-residue connecting peptide 1 insertion (Fragment B) might enhance tryptophan affinity and hence amino acid specificity, because that subdomain constrains the orientation of the specificity helix. We test that hypothesis by characterizing two new constructs: the catalytic domain (fragments A–C) and the Urzyme supplemented with the anticodon-binding domain (fragments A, C, and D). The three constructs, together with the full-length enzyme (fragments A–D), comprise a factorial experiment from which we deduce individual and combined contributions of the two modules to the steady-state kinetics parameters for tryptophan-dependent $^{32}PP_i$ exchange, specificity for tryptophan *versus* tyrosine, and aminoacylation of $tRNA^{Trp}$. Factorial design directly measures the energetic coupling between the two more recent modules in the contemporary enzyme and demonstrates its functionality. Combining the TrpRS Urzyme individually in *cis* with each module affords an analysis of long term evolution of amino acid specificity and tRNA aminoacylation, both essential for expanding the genetic code. Either module significantly enhances tryptophan activation but unexpectedly eliminates amino acid specificity for tryptophan, relative to tyrosine, and significantly reduces tRNA aminoacylation. Exclusive dependence of both enhanced functionalities of full-length TrpRS on interdomain coupling energies between the two new modules argues that independent recruitment of connecting peptide 1 and the anticodon-binding domain during evolutionary development of Urzymes would have entailed significant losses of fitness.

Contemporary aminoacyl-tRNA synthetases (aaRS)[2] comprise three functional modules that implement catalysis, tRNA anticodon recognition, and in some cases hydrolytic editing of misacylated tRNAs. In Class I aaRS, editing is associated with elaboration of a module known as connecting peptide 1 (CP1) (1) that actually divides the catalytic module into two disjoint segments. The kinetic properties of *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase (TrpRS) Urzyme (2, 3) introduced a means to measure the intrinsic and synergistic contributions of these three modular components quantitatively.

Urzyme is a general term introduced to designate an invariant structural core of an enzyme superfamily that has been engineered to restore stability and catalytic activity (2–4). The small size, extensive conservation, and exceptional catalytic activities of Class I and II aminoacyl-tRNA synthetase Urzymes (3) afford presumptive, though perhaps not definitive, evidence for their relationship to ancestral synthetases that have long been extinct. Moreover, the substantial catalytic activities of aaRS Urzymes also afford metrics against which to determine changes induced by restoring deleted modules, individually and in combination.

We report here a full factorial comparison of rate accelerations and amino acid specificity of the TrpRS Urzyme, the full-length enzyme, and the two possible modular combinations that restore either CP1 or the anticodon-binding domain (ABD). The data afford direct, quantitative measurement of a substantial intermodular energetic coupling in a contemporary enzyme. Further, the likely similarity of TrpRS Urzyme to a much earlier ancestral precursor allows us to interpret the results in terms of experimental recapitulation and testing of putative intermediates in the evolutionary development of full-length Class I aaRS from a possible precursor.

Specifically, we test the extent to which enhanced amino acid specificity and tRNA aminoacylation can be attributed to evolutionary recruitment, respectively, of CP1 and the ABD. We find, contrary to that expectation, that both functions necessary for the contemporary enzyme to fully implement the genetic

* This work was supported, in whole or in part, by National Institutes of Health Grants NIGMS 78227 and NIGMS 40906.
[1] To whom correspondence should be addressed: Dept. of Biochemistry and Biophysics, CB #7270, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7263. Tel.: 919-966-3263; Fax: 919-966-2852; E-mail: carter@med.unc.edu.

[2] The abbreviations used are: aaRS, aminoacyl-tRNA synthetase(s); TrpRS, tryptophanyl-tRNA synthetase; CP1, connecting peptide 1; ABD, anticodon-binding domain.

code depend entirely on a ~6-kcal/mol energetic coupling between the two modules.

These results confirm the suggestion (5) that the $\sim 10^4$-fold amino acid specificity in full-length Class I aaRS requires highly cooperative interactions between second sphere amino acid packing in the full catalytic domain that are reinforced by long range communication from the anticodon-binding domain, even in the absence of cognate tRNA. They confirm suggestions elsewhere in the literature (6) that aminoacylation and amino acid specificity are extensively coupled in contemporary full-length aaRS. In a broader context, they constitute a paradigm for the study of long range allosteric interactions between domains in multidomain proteins.

Combining the TrpRS Urzyme in *cis* with each domain also affords a novel factorial analysis of long term evolution of amino acid specificity and tRNA aminoacylation activity. Evolutionary enhancements of the functions necessary for expansion of the genetic code both require interdomain coupling energies that appear to depend on simultaneous growth of both CP1 and the ABD. A perplexing puzzle emerges: if adding either module reduces the ability of evolutionary intermediates to translate the genetic code, then how did contemporary Class I aaRS evolve from ancestral Urzymes?

EXPERIMENTAL PROCEDURES

Construction of TrpRS Modular Structural Intermediates—Restriction enzymes and other molecular biology reagents were obtained from New England Biolabs. The gene for ABC was amplified by PCR from WT TrpRS and that for ACD was ligated from BsaI-digested AC and D fragments. An NdeI restriction site and FLAG tag were included in the primers for the 5′ end. A His_6 tag and HindIII restriction site were added at the 3′ end of both genes. Following double digestion with NdeI and HindIII, the ABC and ACD constructs were cloned to *Escherichia coli* expression plasmid vector pET42 (Novagen). Both constructs were confirmed by DNA sequencing. Bacterial strain BL21(DE3)pLysS (Novagen) was used for the expression of all constructs. In later work, the ABC and ACD fragments were expressed as MBP fusion proteins, which were assayed without cleavage. A limited number of comparisons between steady-state kinetics of the fusion proteins and the peptides themselves showed indistinguishable catalytic properties.

Expression and Purification—The ABC and ACD proteins were expressed and purified using nickel-nitrilotriacetic acid resin according to the manufacturer's instructions (Qiagen). Overnight cultures of *E. coli* transformed with recombinant plasmid were diluted in fresh LB medium with kanamycin and chloramphenicol to $A_{600} = 0.4$ at 37 °C. Then isopropyl β-D-thiogalactopyranoside was added to a final concentration of 1 mM to the culture broth. After 22 h of induction at 18 °C, the cells were harvested by centrifugation at 5,000 × *g* at 4 °C for 20 min. Cell pellets were resuspended in 1/10 volume of chilled lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole, pH 7.4). Resuspended cells were disrupted by sonication on ice and centrifuged at 12,000 × *g* at 4 °C for 30 min. The crude extract was incubated with nickel-nitrilotriacetic acid resin overnight at 4 °C with shaking and then poured into a Kontes Flex-Column. The column was washed with 12 volumes of washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 20 mM imidazole, pH 7.4), and the proteins were eluted with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 300 mM imidazole, pH 7.4). Eluted fractions were pooled and concentrated using an Amicon PM10 Ultra membrane and stored at −20 °C in 40% glycerol.

Active Site Titration—The reaction mixture for the assay consisted of 50 mM HEPES (pH 7.5), 10 mM $MgCl_2$, 100 mM KCl_2, 1 mM DTT, 10 μM γ-^{32}P-labeled ATP, 2.5 mM Trp, 10 units/ml inorganic pyrophosphatase. We added 5 μM enzyme and incubated at 37 °C. At specific time points, the reaction was stopped by adding 3 μl of reaction mixture to 6 μl of quenching solution containing 400 mM sodium acetate (pH 5.0), 0.1% SDS. 2 μl of this mixture was spotted on a prewashed polyethyleneimine cellulose TLC plate (Sigma). To separate ATP from (7) PP_i, we soaked the TLC plates in 750 mM KH_2PO_4 (pH 3.5) and 4 M urea buffer. The TLC plate developed in that buffer was dried, exposed 30 min on an image plate, scanned on the Typhoon Scanner, and quantified with the ImageJ software (8).

The resulting active fractions are comparable to those obtained for the full-length enzyme. In this manner, they address the important question of whether or not the constructs are properly folded. That question is nearly impossible to address more fully without high resolution structural studies, which are outside the scope of this work. Low resolution spectroscopic measurements, such as circular dichroism would add little to the strength of the active fractions, because we have no basis to interpret such spectra in terms of proper or improper folding.

Michaelis-Menten Kinetics—$^{32}PP_i$ exchange assays were done at 37 °C and initiated with 10 μl of enzyme to 190 μl of assay mix: 0.1 M Tris-Cl, 0.01 M KF, 5 mM $MgCl_2$, 2 mM ATP, 2 mM tryptophan, 10 mM 2-mercaptoethanol (pH 8.0) plus 2 mM $^{32}PP_i$ at a specific radioactivity between 1 and 2×10^5 CPM/M. Michaelis-Menten kinetics were examined by varying ATP, tryptophan, and tyrosine concentrations determined by range finding. Assays were performed on 96-well plates, using Whatman filter plates and a Promega Corporation Vacman vacuum manifold for filtrations. All $^{32}PP_i$ exchange assays were processed by eluting [^{32}P]ATP from charcoal with pyridine as described (9). All assays were replicated four times, and multiple assays were performed on different dates to improve the accuracy of estimates made by regression analysis (Table 1). Turnover numbers were corrected for the active factions, which ranged between 0.1 and 0.9.

Aminoacylation was assayed as described (10). Reaction mixtures contained 50 mM HEPES (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 10 mM ATP, 250 μM amino acid, enzyme (5 nM wild type; 1–5 μM Urzyme), and a range of tRNA concentrations (0, 0.5, 1.0, 2.5, 5.0, 10.0, and 15.0 μM). Reactions were run at 37 °C, for variable lengths of time, determined to be in the linear range. Aliquots (1 μl) were removed at varying time points and quenched in 400 mM sodium acetate (pH 5.2). Aminoacylated tRNA was digested with 0.1 mM P1 nuclease (Sigma), spotted on prewashed PEI cellulose TLC plates (Sigma), and developed in 100 mM ammonium acetate, 5% acetic acid. Dried TLC plates were quantified by phosphorimaging analysis. TLC profiles afford quantitative estimates for the amounts of acylated and unacylated A76 bases. The acylated product produced at a given time is thus the product of the

TABLE 1
$^{32}PP_i$ exchange steady-state kinetic parameters for TrpRS and its modular fragments

Construct	ATP		Trp		Tyr	
	k_{cat}	K_m	k_{cat}	K_m	k_{cat}	K_m
	s^{-1}	mM	s^{-1}	mM	s^{-1}	mM
Full-length	5.1 ± 0.44	0.3 ± 0.08	1.7 ± 0.21	0.0033 ± 0.0010	0.04 ± 0.01	0.28 ± 0.13
ABC	0.002 ± 0.0006	0.26 ± 0.06	0.002 ± 0.0007	0.39 ± 0.32	0.003 ± 0.0054	1.8 ± 3.0
ACD	0.0008 ± 0.0003	0.24 ± 0.22	0.002 ± 0.0008	0.058 ± 0.041	0.0022 ± 0.0022	2.8 ± 4.8
Urzyme (AC)	0.0013 ± 0.0018	0.6 ± 0.4	0.0009 ± 0.0006	2.2 ± 1.0	0.000058 ± 0.00002	1.4 ± 1.1

TABLE 2
Steady-state kinetic parameters for $tRNA^{Trp}$ aminoacylation by TrpRS constructs

	k_{cat}	K_m	k_{uncat}, k_{cat}/K_m	Acceleration	$\Delta(\Delta G^{\ddagger})_{rel}$
	s^{-1}	mM			kcal/mol
Uncat			0.00008	1	0
Full-length	0.6 ± 0.21	0.0009 ± 0.0005	700000 ± 140000	8.7E9	−13.5
ABC	0.00043 ± 0.00006	0.005 ± 0.001	90 ± 10	1.10E6	−8.2
ACDwt	0.001 ± 0.0009	0.0055 ± 0.0037	172 ± 36	2.20E6	−8.6
ACD_des	0.00048 ± 0.00006	0.0026 ± 0.00009	187 ± 25	2.30E6	−8.7
Urzyme	0.0017 ± 0.0009	0.0057 ± 0.003	298 ± 4.0	3.70E6	−9.0

initial tRNA concentration times the fraction of acylated A76. Steady-state kinetic parameters (Table 2) were determined using JMP (11) and corrected for the fraction of active enzyme and aminoacylatable tRNA (0.2–0.55).

Factorial Analysis of Interdomain Thermodynamic Coupling—Factorial experimental design (12) is formally equivalent to multivariant thermodynamic cycles (13). Both measure the impact of each perturbation from a set in the context of all other perturbations in the set. Key to the relationship between experimental results and coupling energies is the experimental design matrix. Each experiment corresponds to a row in this matrix, giving the states of each perturbation and their higher order interactions. For the double variant cycle involving the CP1 and ABD domains, the row corresponding to the Urzyme has the form [0,0,0]. The catalytic domain ABC has the form [1,0,0] because it lacks the ABD (0) and has zero also in the third column, which is the product of columns A and B and corresponds to the A*B interaction. The row corresponding to the ACD construct is [0,1,0], and WT TrpRS has [1,1,1]. Reference values of 2.7E-8/s and 8.0E-5/s were subtracted from first rate constants (k_{cat}) for $^{32}PP_i$ exchange and tRNA aminoacylation, to give relative rate accelerations. Specificity ratios, {k_{cat}/K_m (Trp)/k_{cat}/K_m (Tyr)}, were evaluated for all appropriate pairs of measurements of the numerator and denominator.

Without experimental data, contributions of the domains and their interaction energies are unknown. Additional column vectors correspond to experimental values for dependent variables. Free energies ΔGk_{cat}, ΔGK_M, and $\Delta Gk_{cat}/K_m$ are additive and hence can be used to construct linear models from the state columns of the design matrix. These additional column vectors are often prepended to the design matrix. Main and interaction free energies for kinetic parameters and specificity were estimated jointly by regression analysis using JMP (11). Regression models identify the coefficients associated to each column that produce the best least squares fit to the column of dependent variables. These coefficients derive the units of free energy in kcal/mol from the units of the dependent variable columns. We previously have discussed the advantages of this procedure (15), compared with the stepwise procedure described by Horovitz and Fersht (13).

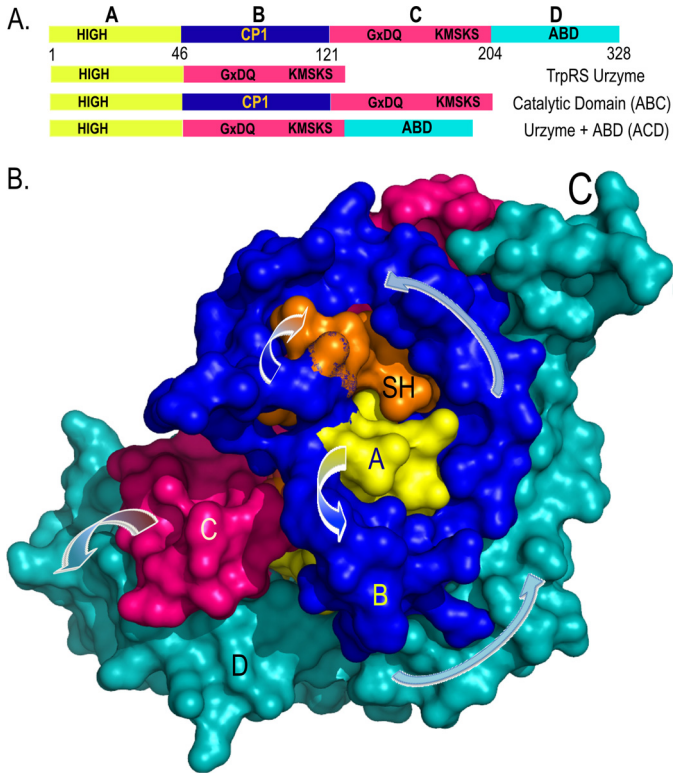

FIGURE 1. **Intermediate TrpRS constructs.** *A*, schematic maps of modular fragments in the TrpRS constructs. *B*, physical interaction of modular components A (*yellow*), B (*blue*), C (*magenta*), and D (*teal*) relative to the specificity-determining helix in the TrpRS monomer. *Arrows* indicate the course of the polypeptide chain, from the N terminus in the center and to the back, to the C terminus in the *upper right*. Both B and C circumvent both sides of the A fragment containing the ATP and tryptophan-binding sites and the specificity helix (*SH*), which forms an essential boundary to the amino acid binding subsite. The dimer interface is at the *upper right* and involves both B and D fragments. Molecular dynamics simulations indicate that the specificity helix rotates away from the A fragment in the absence of tryptophan. The B fragment (CP1) appears to constrain the orientation of the specificity helix, providing significant motivation for this work.

RESULTS

Class I Aminoacyl-tRNA Synthetase Structural Modules—Native *B. stearothermophilus* TrpRS is a dimer composed of 35.6 K_d monomers. TrpRS monomer architecture entails hierarchical levels of modularity (Fig. 1). As a Class I aaRS, it con-

tains two domains that move as rigid bodies in the catalytic cycle of tryptophan activation (17). The two domains correspond roughly to the Rossmann dinucleotide binding fold containing the active site and a C-terminal helical domain containing the anticodon-binding determinants. However, the two catalytic signatures, TIGN and KMSKS, belong to the Rossmann fold yet move with the anticodon-binding domain (18).

The Rossmann fold is itself a mosaic of three modular components: the first and second α-β-α crossover connections (Fig. 1*A*, *fragments A* and *C*) are interrupted by a 76-residue insertion homologous to longer fragments in all Class I aaRS that are commonly termed Connecting Peptide 1 (CP1) (1). Although both contribute to full activity in the intact protein, neither the CP1 subdomain (fragment B) nor the anticodon-binding domain (fragment D) are necessary for catalysis of tryptophan activation by ATP (2, 9).

The N- and C-terminal β-α-β crossover connections of the Rossmann fold are fused in Class I aaRS Urzymes by complete elimination of CP1. The Urzymes also lack the ABD. The resulting 130 residue TrpRS Urzyme accelerates ATP-dependent tryptophan activation $\sim 10^8$-fold and its K_m for ATP is within experimental error of full-length, wild-type enzyme. It accelerates $tRNA^{Trp}$ acylation $\sim 10^6$-fold, and its K_m for $tRNA^{Trp}$ is within an order of magnitude of the wild-type value (3). However, its K_m for tryptophan is >1 mM compared with 2–3 μM in full-length TrpRS. Despite its weak tryptophan binding, the TrpRS Urzyme has even less activity for the related amino acid, tyrosine, and the specificity ratio redetermined here, $(k_{cat}/K_m)_{Trp}/(k_{cat}/K_m)_{Tyr}$ is ~10.

Whereas Class I Urzymes and the anticodon-binding domain are relatively unambiguous, the CP1 insertions vary widely in different Class I synthetases. The 76-residue TrpRS CP1 insertion forms an annulus surrounding the amino acid binding pocket (2). Thus, it cannot serve the hydrolytic editing function of more extensive CP1 domains in Class IA aaRS for valine, leucine, and isoleucine. As illustrated schematically in Fig. 1, CP1 (*fragment B*) circumscribes both the A-fragment and the N terminus of the specificity helix, residues 125–136 in fragment C, forming an exoskeleton constraining the orientation of that helix.

Molecular Dynamics simulations of the TrpRS Urzyme with Mg^{2+}·ATP alone showed that the specificity helix, containing many of the side chains interacting with bound tryptophan, reorients significantly during the course of a 70-ns simulation (2). However, this helix remains properly oriented in the presence of tryptophan. These observations motivated our earlier suggestion that the TrpRS CP1 insertion might enhance specific amino acid binding intrinsically, by acting as an exoskeleton to constrain the orientation of the specificity helix, residues 125–136 (WT numbering). Our thought was that orientational constraint by CP1 might reduce the unfavorable entropy change of reorienting the helix on tryptophan binding, increasing affinity, and perhaps also enhancing amino acid specificity.

To test this hypothesis, we expressed two new TrpRS constructs, ABC and ACD (Fig. 2), each including the Urzyme plus one of the two modules missing from the full-length enzyme. Both constructs show improved catalytic acceleration of tryptophan activation in the $^{32}PP_i$ exchange reaction. The presence

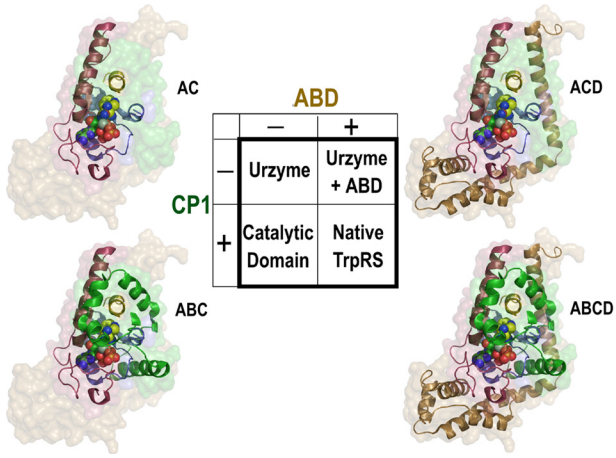

FIGURE 2. **Schematic illustration of TrpRS modularities.** Intermediate constructs characterized in this paper, defined schematically in Fig. 1*A* and compared with the crystallographic structure of the pre-transition state TrpRS dimer (Protein Data Bank code 1mau). Note that although both B and D fragments contribute surface area to the dimer interface, the core of the interface is provided largely by the B fragment. All graphics are based on crystal structures of full-length *B. stearothermophilus* TrpRS.

of the CP1 insertion decreases K_m ~6-fold, as expected. However, the anticodon-binding domain also enhances tryptophan affinity, albeit to a lesser extent. Moreover, adding back either the CP1 insertion or the anticodon-binding domain actually eliminates the preferential tryptophan activation seen in the Urzyme and reduces the aminoacylation activity of the Urzyme by a comparable amount.

Comparison of the ABC and ACD constructs with each other and with the full-length TrpRS and its Urzyme is confounded by the fact that the Urzyme itself is unstable without compensating mutations made to the newly exposed surface area created by removing CP1 and the anticodon-binding domain. For the present purposes, we constructed the ABC catalytic domain using the native *B. stearothermophilus* sequence. However, to assess the contributions of these mutations, we added the C-terminal D fragment containing the anticodon-binding determinants to both wild-type and redesigned AC Urzyme (2) to give two different ACD constructs, ACDdes and ACDwt. These two constructs behave similarly but may differ in ways whose statistical significance cannot be demonstrated without additional experiments.

The similar functionalities of the ACDwt and ACDdes hybrid both justifies using the redesigned Urzyme and opens the possibility of investigating the interaction between the Urzyme and ABD more deeply by restoring wild-type residues Val-13, Ile-14, Ile-16, Tyr-19, Leu-23, and Ile-204, which are located in the interface between the two domains. The complementary experimental construct for the ABC catalytic domain would entail using Rosetta to redesign the CP1 insertion to complement mutations between it and the redesigned Urzyme. Both experiments are now feasible and under consideration.

Histograms in Figs. 3–7 summarize the double-variant thermodynamic cycles associated with adding either or both of the two domains to the Urzyme. These data (Table 1) constrain the evolution from TrpRS Urzyme to the contemporary enzyme in unexpected ways. With few exceptions (K_m for tyrosine), the regression models from which they were obtained have uni-

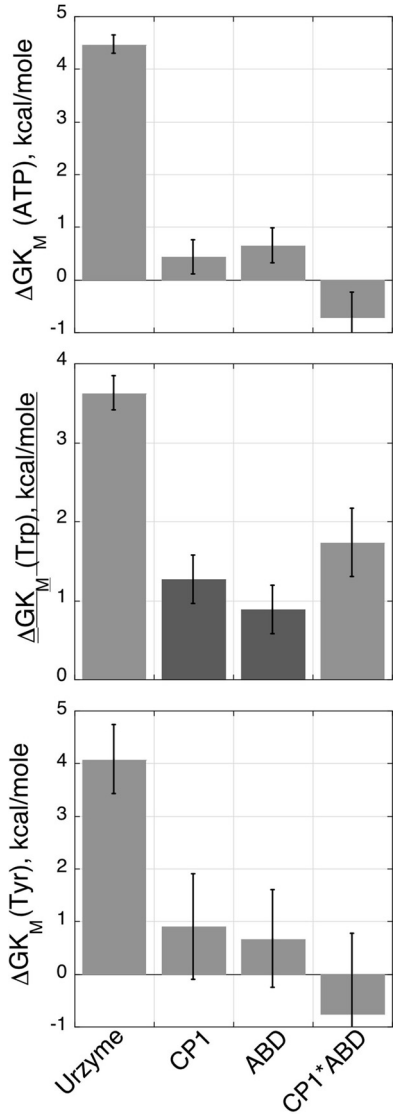

FIGURE 3. **Modular contributions to ground state substrate affinity.** Adding either the CP1 or ABD domains to the Urzyme increases affinity for all three substrates, ATP, tryptophan, and tyrosine, but do so preferentially for tryptophan (*dark bars*). The *error bars* here and in Figs. 4–7 are standard errors assigned by regression modeling. The ratio of the *error bar* to the histogram height is representative of the Student's *t* test values (available from the author on request) and hence is related inversely to the *t* test probability under the null hypothesis.

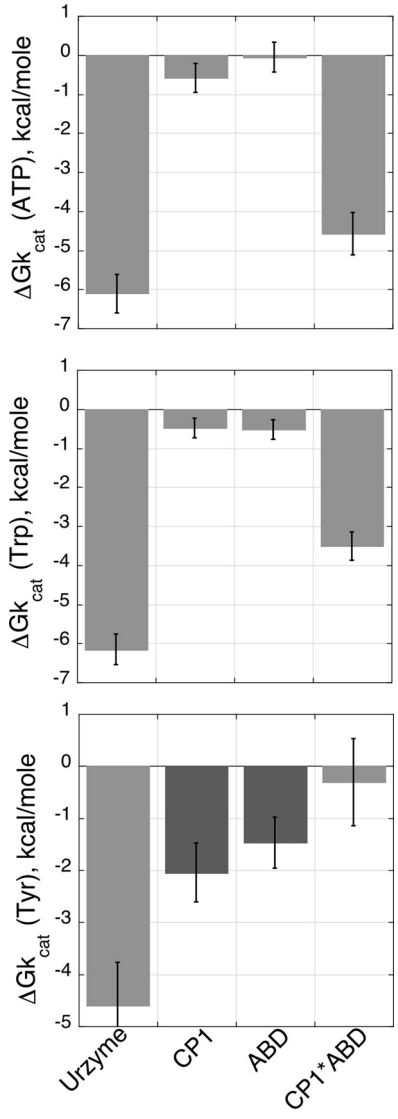

FIGURE 4. **Modular contributions to transition state affinities, (ΔGk_{cat}).** The *dark bars* emphasize the preferential acceleration of tyrosine activation by either CP1 or the ABD.

formly high R^2 values and hence the relative domain contributions are unambiguous. Because Michaelis-Menten parameters for the $^{32}PP_i$ exchange assay represent ground state (K_m) and transition state (k_{cat}) affinity for the relevant substrate (19, 20), the histograms represent free energy contributions in kcal/mol afforded by the Urzyme, the addition of either domain, and the domain-domain interaction energy to the behavior of the contemporary enzyme. The latter is the sum of the four bars in each histogram.

CP1 and ABD Alone Both Preferentially Increase Amino Acid Affinity—Increases in affinity are induced by both domains for all three substrates, as indicated by positive ΔGK_M values (Fig. 3). As expected, the addition of CP1 increases affinity of the catalytic domain for tryptophan (bold histograms) and actually induces higher affinity for tryptophan than does the ABD. The two-way interaction reduces ATP and tyrosine affinity, while increasing tryptophan affinity substantially.

CP1 and ABD Alone Have Little Effect on TS Binding of Either Tryptophan or ATP but Increase That of Tyrosine—Histograms for ΔGk_{cat} are notable in that neither individual domain significantly impacts transition state affinity for either of the two appropriate substrates, whereas both domains substantially improve transition state affinity for the noncognate amino acid, tyrosine (Fig. 4). Notably, whereas the two-way interaction between CP1 and ABD domains stabilizes the TS ATP and tryptophan complexes substantially (3–5 kcal/mol), it has essentially no effect on the TS affinity of the noncognate tyrosine complex.

CP1 and ABD Both Increase Proficiency of Amino Acid Activation—Both domains individually enhance the apparent second order rate constant, $\Delta Gk_{cat}/K_m$, primarily through increased ground state affinity (Fig. 5). The CP1 insertion consistently shows a slightly greater effect than the ABD. Curiously, this effect is most pronounced for the noncognate amino acid, approaching −3 kcal/mol mainly because of improved TS

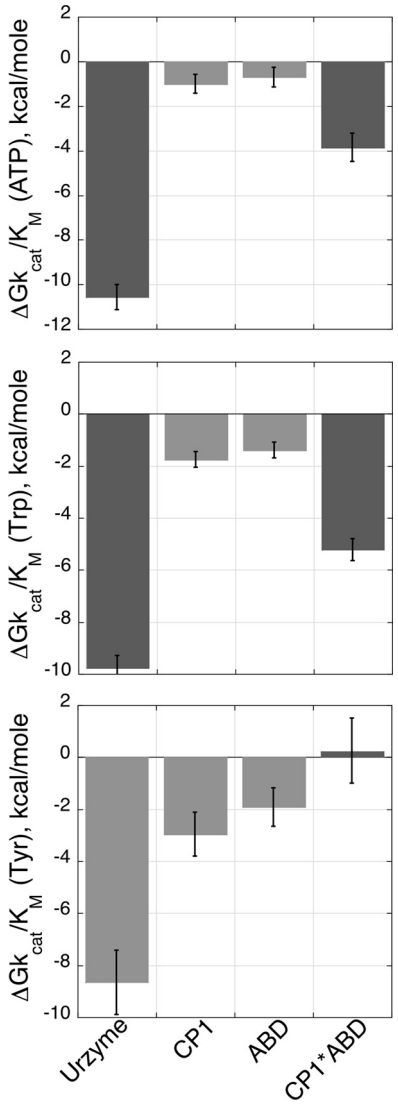

FIGURE 5. **Modular contributions to overall catalytic proficiencies, ($\Delta Gk_{cat}/K_m$).** The *dark bars* emphasize that the Urzyme and native enzyme contribute essentially the total rate acceleration for the apparent second order rate constant.

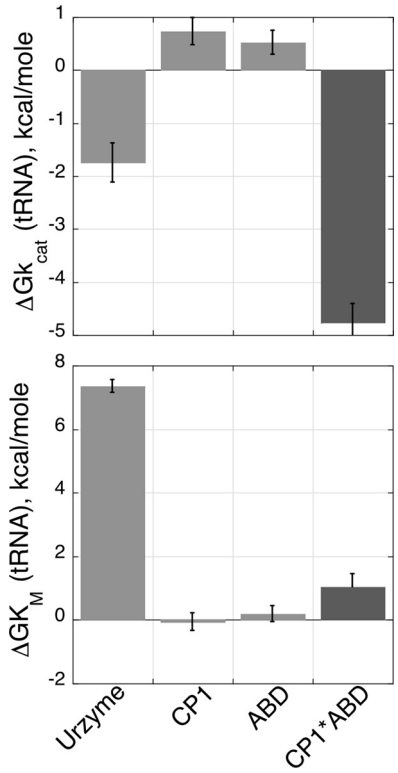

FIGURE 6. **Modular contributions to Michaelis-Menten parameters for tRNA aminoacylation.** The *dark bars* emphasize the magnitude of the contributions made by the CP1*ABD interaction energy to catalysis (large) and substrate affinity (modest).

stabilization. We will see in a subsequent section that this effect of CP1 abolishes the ~10-fold relative specificity of the Urzyme for tryptophan *versus* tyrosine. As for the other kinetic parameters, the ABD has an effect that approaches that exhibited by CP1, enhancing $\Delta Gk_{cat}/K_m$ by nearly 2 kcal/mol. The most remarkable effects on proficiency are those of the two-way interaction, which contributes 4–5 kcal/mol to proficiency with both ATP and tryptophan but decreases it for tyrosine.

TrpRS Urzyme Accounts for Major Fractions of WT Substrate Affinity and TS Stabilization—We previously reported (2, 3) that the invariant cores that compose the Class I and II Urzymes exhibit the major portion of the catalytic free energies exhibited by the full-length, native enzymes (Figs. 3–5). The additional data shown here reinforce that conclusion. Indeed, the relatively recent CP1 and ABD domains in general induce modest, nonspecific improvements in the basal activity of the TrpRS Urzyme unless they are both present, in which case their impact is functionally decisive, although still smaller than the catalytic contribution of the Urzyme itself.

tRNA Aminoacylation Requires Similar Multiple Modular Interactions—The TrpRS Urzyme accelerates $tRNA^{Trp}$ aminoacylation by 10^6-fold (2, 3) (Fig. 6). Adding either CP1 or the anticodon-binding domain reduces this acceleration by small but significant amounts because the increase in ground state binding affinity cannot overcome the reduction in transition state stabilization by the intermediate constructs. The aminoacylation rate of the Urzyme is exceeded only by full-length TrpRS where increased ground and transition state affinities work together.

Improved Ability of Contemporary TrpRS to Translate the Genetic Code Emerges Only from Synergy of Both CP1 and ABD Modules—Interaction between the two modules is crucial for amino acid specificity (Fig. 7). Whereas the Urzyme exhibits ~10-fold specificity for tryptophan, neither the ABC nor the ACD construct discriminates between the two similar substrates. This result implies profound energetic coupling ($\Delta\Delta G_{int} \sim -6$ kcal/mol) between the CP1 subdomain and the anticodon-binding domain in suppressing tyrosine activation (Table 3). Quantitatively similar coupling (~−5.6 kcal/mol) enhances the tRNA aminoacylation rate in full-length TrpRS.

DISCUSSION

Conventional analyses of enzyme function perturb individual residues by mutagenesis. Modular engineering according to a factorial design measures the strength of long range interdomain coupling energies directly, by experimental manipulations of entire blocks of genetic information corresponding to secondary or supersecondary structures. Intrinsic and coupled

energetic contributions of two modules deleted from the TrpRS Urzyme to substrate affinity, catalysis, and specificity are relevant to understanding both the mechanistic behavior of contemporary enzymes and how that behavior evolved.

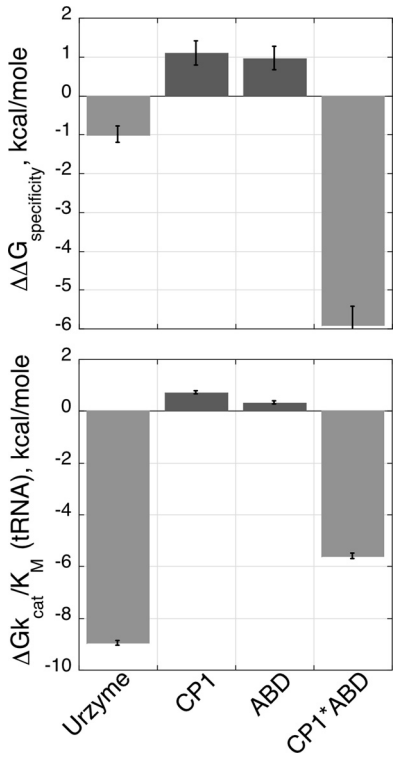

FIGURE 7. **Modular contributions to translating the genetic code.** The small magnitude of the *dark bars* emphasizes that amino acid specificity and tRNA aminoacylation both depend entirely on cooperativity between CP1, the anticodon-binding domain, and the active site. $\Delta\Delta G_{specificity}$ is for specificity ratios, and $\Delta Gk_{cat}/K_m$ is for $tRNA^{Trp}$ aminoacylation, which display nearly identical modular profiles. Specifically, addition of either intermediate module (CP1 or anticodon-binding domain) results in extensive loss of specific rejection of tyrosine, whereas the intact enzyme has a specificity ratio of ~4,000.

Domain-Domain Energetic Coupling Accounts Fully for the Evolutionary Improvements in Specificity and tRNA Aminoacylation by Contemporary TrpRS

Emergence of the full Class I aaRS superfamily necessary for full expression of the genetic code required substantial increases in selective amino acid affinity. That enhanced specificity, in turn, requires quite substantial intermodule communication.

Previous efforts to alter the amino acid specificity of TrpRS and other Class I aaRS by combining mutant residues in the immediate vicinity of the amino acid in both TrpRS (21) and GlnRS (22) anticipated the importance of intermodular coupling. Locally directed mutagenesis failed to alter specificity in both cases, whereas considerably more radical engineering of second sphere residues in GlnRS eventually improved specificity for glutamate (5).

Energetic coupling between CP1 and the ABD (−5.9 kcal/mol; Fig. 7; Table 3) is crucial to the specific recognition of tryptophan *versus* tyrosine by full-length TrpRS (−4.9 kcal/mol), especially in view of the reduced specificity (both ~+1.0 kcal/mol) of the two intermediate constructs. The specificity of full-length TrpRS is therefore entirely a consequence of the cooperative action of the two more recent modules. Quantitatively similar arguments apply for the rate acceleration of tRNA aminoacylation.

The magnitudes contributed by interdomain coupling to specificity for tryptophan *versus* tyrosine and transition state stabilization for aminoacylation (−5.6 kcal/mol) are remarkably similar to the value (−6 kcal/mol) (15) previously determined for the allosteric contribution of the five-way interaction between four residues in the D1 switch (Ile-4, Phe-26, Tyr-33, and Phe-37) and Mg^{2+} to transition state stabilization during tryptophan activation. The four D1 switch residues undergo the most significant side chain repacking during domain motions observed from *B. stearothermophilus* TrpRS crystal structures

TABLE 3

Illustrative factorial analyses of modular contributions to ground and transition state affinities, amino acid specificity, and tRNA aminoacylation as functions of the presence/absence of CP1 and the anticodon-binding domain

ΔGk_{cat} and ΔGK_m values were calculated from parameters in Tables 1 and 2 using $\Delta\Delta G = -RT\ln(k_{cat}, K_m)$. Regression models were estimated using JMP (11) as described under "Experimental Procedures" (15, 36).

Factor	Coefficient	Standard error	Student's *t* test	*P*(*t*)
	kcal/mol	*kcal/mol*		
ΔGk_{cat} (from ATP dependence)				
Intercept	10.4	0.45	23.22	<0.0001
Urzyme	−6.12	0.49	−12.47	<0.001
CP1	−0.59	0.37	−1.57	0.15
ABD	−0.65	0.37	−0.17	0.87
CP1*ABD	−4.59	0.54	−32.36	<0.0001
ΔGK_m (Trp)				
Urzyme	3.62	0.22	17.73	<0.0001
CP1	1.26	0.31	4.12	0.0014
ABD	0.88	0.31	2.88	0.014
CP1*ABD	1.73	0.43	3.99	0.0018
$\Delta G_{specificity}$ (Trp *vs.* Tyr)				
Urzyme	−1.01	0.21	−4.76	<0.0001
CP1	1.09	0.32	3.42	0.0013
ABD	0.96	0.29	3.24	0.0022
CP1*ABD	−5.91	0.48	−12.4	<0.0001
$\Delta Gk_{cat}/K_m$ (tRNA aminoacylation)				
Intercept	5.60	0.08	67.44	<0.0001
Urzyme	−8.97	0.10	−93.59	<0.0001
CP1	0.71	0.07	10.52	<0.0001
ABD	0.31	0.06	5.24	0.0004
CP1*ABD	−5.61	0.10	−58.53	<0.0001

(23). We concluded that the catalytic assist by Mg^{2+} arose from coupling to domain motion.

ABDs often but not always bind to the anticodon of the cognate tRNA. LeuRS must acylate a large number of isoaccepting tRNAs, precluding specific recognition of the anticodon itself. It is noteworthy that the ABD appears similarly to be tightly coupled to CP1 in LeuRSs, as evidenced by linked movements during aminoacylation and editing of misacylated $tRNA^{Leu}$ (24–27).

The present data therefore shed substantial new light on the central functional importance of domain motion in contemporary TrpRS and by implication other Class I aaRS. Much previous work hinted at the importance of interdomain coupling in aaRS (Refs. 27 and 28; reviewed in Ref. 29; and observed originally in Ref. 30) without, however, estimating either its magnitude or the purpose(s) served by such effects. Our work affords a direct determination of both effects: evolutionary adaptation has distributed the energetic sources of major enhancements (*i.e.*, ~6 kcal/mol) in three primary functions into coupling between domains.

Molecular Evolution: Recapitulating the Enhancement of Catalytic Activity and Substrate Specificity by Modular Additions

We describe here a new way to examine how catalytic activity and substrate specificity, the basic functions of enzymes, develop as an enzyme family evolves by accumulating new modules.

We reconstructed putative events in the molecular evolution of the Class I aminoacyl-tRNA synthetase superfamily. The TrpRS Urzyme represents the most highly conserved elements of the superfamily and by Ockham's razor probably closely resembles an ancestral form dating to the era in which the genetic code first began to assemble. The Urzyme lacks both the catalytic activity and the full amino acid specificity of contemporary TrpRSs. We show here that its CP1 and ABD modules are both necessary to account for the adaptive enhancements of fitness in the contemporary enzyme. They are also, obviously, sufficient.

We began with the TrpRS Urzyme, a simple, 130-residue polypeptide representing the most highly conserved architectural elements (2, 9). The Urzyme is both less active than modern TrpRS and less specific for its cognate amino acid. Here, we recapitulate two limiting constructions to answer the question: which functions, and in what proportions, are enhanced by restoring the modules, one at a time, that were deleted to make the Urzyme.

Our results yield the following three conclusions:

All Four TrpRS Constructs Interact Equally Well with ATP—K_m(ATP) values for the AC, ABC, and ABD constructs (3.8E-4 ± 7.5E-5) are within experimental error of those observed for full-length TrpRS (3.0E-4 ± 7.9E-5). This conclusion is consistent with the fact that neither CP1 nor the ABD make ground state interactions with ATP. One residue from CP1 that interacts with the γ-phosphate group (Lys-111) contributes only to the transition state affinity (36). This lysine accounts for −1.62 kcal/mol and 35% of the −4.6 kcal/mol contributed by the CP1*ABD interaction but contributes little to the activity of the catalytic domain itself (Table 3).

The near invariance of ATP binding affinity and progressive increases in k_{cat} in all four constructs argue that the initial function of modules present in the Urzyme was to bind and mobilize the chemical free energy stored in ATP and that this functionality readily lent itself to amino acid activation. ATP binds initially to the TIGN sequence in the N-terminal crossover connection without making contact with other residues in the Rossmann fold. This 46-residue module also contains important switching residues (Ile-4, Phe-26, Tyr-33, and Phe-37) that are involved in allosteric behavior (15). The central importance of NTP utilization in biology suggests that these factors may account for the fact that this module is arguably the most highly conserved packing motif in the proteome (31).

Both CP1 and the Anticodon-binding Domain Enhance Amino Acid Affinity—In marked contrast to their similar ATP affinity, the four K_m (Trp) values (Table 1) range over approximately 3 orders of magnitude. As we suggested, the presence of CP1 reduces K_m (Trp) 6-fold (0.0022 to 0.00039 M). Unexpectedly, adding the anticodon-binding domain alone has a similar effect (0.0022 to 0.00058 M; 4-fold). Neither of these effects approaches the magnitude of what is observed in the full-length enzyme (~500-fold).

Functionally Beneficial Effects of CP1 and the ABD Arise Exclusively via Energetic Coupling between Them—Figs. 3–7 document that neither module can function productively without interacting with the other. The presence of either module alone substantially decreases the functionality present in the Urzyme, by eliminating specificity for tryptophan and reducing catalysis of tRNA aminoacylation, which thus arise entirely from synergy between the two modules.

How Do Beneficial Allosteric Interactions Evolve?

The unambiguous conclusion of the three previous sections is surprising and somewhat unsettling: each of the restored modules degrades the fitness of the Urzyme unless they can work together. The observations described in the previous three sections raise an important new question: how did evolutionary recruitment of the two additional modules enhance functionality if the putative intermediates are less functional than the ancestral Urzyme?

Our modular factorial analysis highlights this evolutionary question, without resolving it. One possibility is that Urzymes enabled translation of messages according to a rudimentary code that allowed the elaboration of various stand-alone globular domains that could act in *trans* to enhance specific aminoacylation by intermediate synthetase constructs into which one of the domains had been inserted, before they were recruited into aaRS genes to produce the modern sets of roughly 10 aaRS in each class.

By this hypothesis, allosteric behavior began to evolve prior to the assembly of additional domains into a single gene. Our previous work (15, 31) identified a key component of allosteric behavior in the D1 switch that is contained entirely within the Urzyme. That switching mechanism could have been sensitive to the presence of other modules in *trans*. Which of the two

modules would have been more likely to have been the first to be joined to the Urzyme gene?

It is especially interesting that the CP1 domain cannot promote additional specificity without energetic coupling to the anticodon-binding domain. The TrpRS CP1 domain is unlikely to be a vestige remaining from a larger, free standing editing domain that functioned in *trans*. The Ybak stand-alone editing domain is a globular protein of 151 amino acids (14), and the Class Ia editing domains have ~190 amino acids. No TrpRSs have editing functions, because their CP1 insertions (~75 amino acids) are far too short to carry out hydrolytic editing. Further, they neatly circumscribe the Urzyme itself (Fig. 1*B*). The component helices of TrpRS CP1 start and end at the same location. Moreover, they are conserved elements common to all CP1 insertions and hence likely to be ancestral to the larger editing domains. Hence, they seem unlikely intermediates in stepwise completion of the Class I catalytic domain. Scenarios in which stand-alone editing domains provided temporary support for enhancements afforded by the ABD are therefore unlikely.

The structural homology of the TrpRS CP1 insertion to part of the bridge connecting the active sites to the larger, Class Ia CP1 editing domains suggested that they evolved via a small initial insertion into the ancestral Urzyme (9). However, because the CP1 domain alone actually decreases specificity, it is difficult to imagine that this event preceded addition of the ABD (see Fig. 5 in Ref. 9).

Whereas direct evidence exists for stand-alone editing domains associated with Class II, but not Class I aaRS (32–34), no free-standing anticodon-binding domains have been identified. That they may have existed seems reasonable in light of the idiosyncrasy and multiple usage of the α-helix bundles (most Class I aaR), oligonucleotide binding folds in Class IIB aaRS (35), and Rossmannoid domains Class IIA (29, 31) as ABDs. Thus, it seems more likely that the Urzyme accumulated the rudimentary CP1 domain present in TrpRS in a single event, possibly via a transposable element that eventually invaded all Class I Urzymes and whose selective advantage may have been to enhance amino acid specificity when potentiated by stand-alone ABDs.

The co-evolution of cooperative behavior between the two modules raises a substantive new and potentially general question. Our results highlight the subtlety with which communication developed between the two modules considered here and the primordial catalyst represented by the TrpRS Urzyme. Moreover, they introduce a new experimental approach for further study. Combined with the novel use of high order combinatorial mutagenesis (15) and simultaneous network and co-evolution analysis (27), they offer possible avenues toward deeper understanding.

REFERENCES

1. Burbaum, J. J., and Schimmel, P. (1991) Structural relationships and the classification of aminoacyl-tRNA synthetases. *J. Biol. Chem.* **266,** 16965–16968
2. Pham, Y., Kuhlman, B., Butterfoss, G. L., Hu, H., Weinreb, V., and Carter, C. W., Jr. (2010) Tryptophanyl-tRNA synthetase Urzyme. A model to recapitulate molecular evolution and investigate intramolecular complementation. *J. Biol. Chem.* **285,** 38590–38601
3. Li, L., Francklyn, C., and Carter, C. W., Jr. (2013) Aminoacylating Urzymes challenge the RNA world hypothesis. *J. Biol. Chem.* **288,** 26856–26863
4. Li, L., Weinreb, V., Francklyn, C., and Carter, C. W., Jr. (2011) Histidyl-tRNA synthetase Urzymes. Class I and II aminoacyl tRNA synthetase Urzymes have comparable catalytic activities for cognate amino acid activation. *J. Biol. Chem.* **286,** 10387–10395
5. Corigliano, E. M., and Perona, J. J. (2009) Architectural underpinnings of the genetic code for glutamine. *Biochemistry* **48,** 676–687
6. Ibba, M., and Söll, D. (2004) Aminoacyl-tRNAs. Setting the limits of the genetic code. *Genes Dev.* **18,** 731–738
7. Ledoux, S., and Uhlenbeck, O. (2008) [3′-^{32}P]-Labeling tRNA with nucleotidyltransferase for assaying aminoacylation and peptide bond formation. *Methods* **44,** 74–80
8. Uter, N. T., Gruic-Sovulj, I., and Perona, J. J. (2005) Amino acid-dependent transfer RNA affinity in a class I aminoacyl-tRNA synthetase. *J. Biol. Chem.* **280,** 23966–23977
9. Pham, Y., Li, L., Kim, A., Erdogan, O., Weinreb, V., Butterfoss, G. L., Kuhlman, B., and Carter, C. W., Jr. (2007) A minimal TrpRS catalytic domain supports sense/antisense ancestry of class I and II aminoacyl-tRNA synthetases. *Mol. Cell* **25,** 851–862
10. Francklyn, C. S., First, E. A., Perona, J. J., and Hou, Y.-M. (2008) Methods for kinetic and thermodynamic analysis of aminoacyl-tRNA synthetases. *Methods* **44,** 100–118
11. SAS (2004) *JMP Statistics and Graphics Guide V*, 6 Ed., SAS Institute, Cary, NC
12. Carter, C. W., Jr. (1990) Efficient factorial designs and the analysis of macromolecular crystal growth conditions. *Methods Companion Methods Enzymol.* **1,** 12–24
13. Horovitz, A., and Fersht, A. R. (1990) Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. *J. Mol. Biol.* **214,** 613–617
14. Wong, F. C., Beuning, P. J., Silvers C, Musier-Forsyth K (2003) An isolated class II aminoacyl-tRNA synthetase insertion domain is functional in amino acid editing. *J. Biol. Chem.* **278,** 52857–52864
15. Weinreb, V., Li, L., and Carter, C. W., Jr. (2012) A master switch couples Mg^{2+}-assisted catalysis to domain motion in *B. stearothermophilus* tryptophanyl-tRNA Synthetase. *Structure* **20,** 128–138
16. Deleted in proof
17. Laowanapiban, P., Kapustina, M., Vonrhein, C., Delarue, M., Koehl, P., and Carter, C. W., Jr. (2009) Independent saturation of three TrpRS subsites generates a partially assembled state similar to those observed in molecular simulations. *Proc Nat. Acad. Sci. U.S.A.* **106,** 1790–1795
18. Ilyin, V. A., Temple, B., Hu, M., Li, G., Yin, Y., Vachette, P., and Carter, C. W., Jr. (2000) 2.9 A crystal structure of ligand-free tryptophanyl-tRNA synthetase. Domain movements fragment the adenine nucleotide binding site. *Protein Sci.* **9,** 218–231
19. Cleland, W. W. (1970) Steady state kinetics. in *The Enzymes* (Boyer, P., ed) Vol. II, pp. 1–65, Academic Press, New York
20. Cleland, W. W., and Northrop, D. B. (1999) Energetics of substrate binding, catalysis, and product release. *Methods Enzymol.* **308,** 3–27
21. Praetorius-Ibba, M., Stange-Thomann, N., Kitabatake, M., Ali, K., Söll, I., Carter, C. W., Jr., Ibba, M., and Söll, D. (2000) Ancient adaptation of the active site of tryptophanyl-tRNA synthetase for tryptophan binding. *Biochemistry* **39,** 13136–13143
22. Bullock, T. L., Uter, N., Nissan, T. A., and Perona, J. J. (2003) Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J. Mol. Biol.* **328,** 395–408
23. Kapustina, M., Weinreb, V., Li, L., Kuhlman, B., and Carter, C. W., Jr. (2007) A conformational transition state accompanies tryptophan activation by *B. stearothermophilus* tryptophanyl-tRNA synthetase. *Structure* **15,** 1272–1284
24. Chopra, S., Palencia, A., Virus, C., Tripathy, A., Temple, B. R., Velazquez-Campoy, A., Cusack, S., and Reader, J. S. (2013) Plant tumour biocontrol agent employs a tRNA-dependent mechanism to inhibit leucyl-tRNA synthetase. *Nat. Commun.* **4,** 1417
25. Lincecum, T. L., Jr., Tukalo, M., Yaremchuk, A., Mursinna, R. S., Williams, A. M., Sproat, B. S., Van Den Eynde, W., Link, A., VanCalenbergh, S., Grøtli, M., Martinis, S. A., and Cusack, S. (2003) Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. *Mol. Cell* **11,** 951–963

26. Palencia, A., Crépin, T., Vu, M. T., Lincecum, T. L., Jr., Martinis, S. A., and Cusack, S. (2012) Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* **19,** 677–684
27. Weimer, K. M., Shane, B. L., Brunetto, M.,Bhattacharyya, S., and Hati, S. (2009) Evolutionary basis for the coupled-domain motions in *Thermus thermophilus* leucyl-tRNA synthetase. *J. Biol. Chem.* **284,** 10088–10099
28. Johnson, J. M., Sanford, B. L., Strom, A. M., Tadayon, S. N., Lehman, B. P., Zirbes, A. M., Bhattacharyya, S., Musier-Forsyth, K., and Hati, S. (2013) Multiple pathways promote dynamical coupling between catalytic domains in *Escherichia coli* prolyl-tRNA synthetase. *Biochemistry* **52,** 4399–4412
29. Perona, J. J., and Hadd, A. (2012) Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry* **51,** 8705–8729
30. Jahn, M., Rogers, M. J., and Söll, D. (1991) Anticodon and acceptor stem nucleotides in tRNA(Gln) are major recognition elements for *E. coli* glutaminyl-tRNA synthetase. *Nature* **352,** 258–260
31. Cammer, S., and Carter, C. W., Jr. (2010) Six Rossmannoid folds, including the Class I aminoacyl-tRNA synthetases, share a partial core with the anti-codon-binding domain of a Class II aminoacyl-tRNA synthetase. *Bioinformatics* **26,** 709–714
32. Ahel, I., Korencic, D., Ibba, M., and Söll, D. (2003) Trans-editing of mischarged tRNAs. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 15422–15427
33. An, S., and Musier-Forsyth, K. (2005) Cys-tRNA(Pro) editing by *Haemophilus influenzae* YbaK via a novel synthetase·YbaK·tRNA ternary complex. *J. Biol. Chem.* **280,** 34465–34472
34. Perona, J. J., and Gruic-Sovulj, I. (2013) Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. *Top. Curr. Chem.* 10.1007/128_2013_456
35. Arcus, V. (2002) OB-fold domains. A snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.* **12,** 794–801
36. Weinreb, V., Li, L., Campbell, C. L., Kaguni, L. S., and Carter, C. W., Jr. (2009) Mg^{2+}-assisted catalysis by *B. stearothermophilus* TrpRS is promoted by allosteric effects. *Structure* **17,** 952–964